

## Comment

# The use of family health services authority registers as a sampling frame in the UK: a review of theory and practice

H R Roberts, L Rushton, K R Muir, R Dengler, C A C Coupland, C M Jenkinson, A Ruffell, C E D Chilvers

In epidemiological and social scientific studies the choice of an appropriate, accurate sampling frame listing eligible subjects is a methodological decision which will influence the representativeness of the sample selected, and hence the generalisability of findings. The most widely used population sampling frame of adult individuals in the UK is the electoral roll.<sup>1,2</sup> Telephone listings have become more popular in recent years while other studies use sampling frames related to property by which to identify individuals.<sup>3,4</sup>

An increasingly used population sampling frame is the list of persons registered with UK general practitioners (GP) held by family health services authorities (FHSA). These registers were developed so that FHSAs (previously Family Practitioner Committees and Executive Councils) could arrange payment of GPs for National Health Service patients.<sup>1</sup> FHSAs are usually based on counties or metropolitan boroughs, and are accountable to regional health authorities and so to the NHS Executive. Recent major developments in FHSA registers mean that they have a number of important advantages over more usual sampling frames.

- Everyone registered with a GP is registered with an FHSA. Approximately 95% of the population is included.<sup>5</sup>
- Data are computerised and so relatively easy to manipulate. Patients may be readily stratified by one or more characteristics. Computerised searches may be made, for example, by surname to identify some ethnic groups.<sup>6</sup>
- Registers routinely contain three demographic details of use to researchers: gender, age, and postcode.
- They provide doctors' reference numbers and patients' NHS numbers. Thus, patients may be linked to other health service data.
- Lists are updated regularly, although there are inevitable delays in adding new patients and removing those no longer eligible.
- Sponsorship of a survey by an FHSA may increase survey response.<sup>7</sup>

This review reports our experience in using FHSA registers as a sampling frame in three different types of epidemiological and social

scientific studies. We draw some general conclusions on the characteristics of FHSA registers which may be of use to others contemplating their use. We also comment on future developments in the use of these registers.

### Summaries of three studies using FHSA registers

Nottingham University's Department of Public Health Medicine and Epidemiology is undertaking three major projects that utilise FHSA registers: a region-wide population "lifestyle" survey; the Trent region's component of a national case-control study of childhood cancer, and a randomised controlled trial of prevention of osteoporosis. In this section each study is described briefly, emphasising the role of FHSA registers, before reviewing collective experiences in approaching FHSAs.

#### STUDY A: TRENT HEALTH LIFESTYLE SURVEY (THLS)

The THLS is a series of cross sectional population surveys (1991-1994). The study is funded by Trent Regional Health Authority and collects information which is essential to its strategic policies, planning, and practice in health promotion across the region. Data are collected by post, using a 16 page, self completion questionnaire.<sup>8</sup>

In 1992, all eight FHSAs within Trent used their registers to provide samples of 1800 adults for each of Trent's 12 district health authorities. Samples were stratified by gender and, in all but one case, age (16-34, 35-54 and 55-70 years) before systematic selection of samples using software developed by the Family Health Services (FHS) Computer Unit at Exeter. The age, gender, forename, surname, address, and postcode of individuals in the sample were down-loaded to ASCII files.

#### STUDY B: UNITED KINGDOM CHILDHOOD CANCER STUDY (UKCCS)

The UKCCS is an interview based, case-control study which aims to identify causative

Department of  
Public Health  
Medicine  
and Epidemiology,  
University of  
Nottingham  
Medical School,  
Nottingham  
H R Roberts  
L Rushton  
K R Muir  
R Dengler  
C A C Coupland  
C M Jenkinson  
C E D Chilvers

Nottinghamshire  
Family Health  
Service Authority  
A Ruffell

Correspondence to:  
Professor C E D Chilvers.

Accepted for publication  
March 1995

factors in childhood cancer. The study is organised on a regional basis by the United Kingdom Co-ordinating Committee on Cancer Research (UKCCCR), comprising the major cancer charities and the Medical Research Council. The study is funded by the UKCCCR and directly by its members.

Cases are identified primarily through consultant paediatric oncologists. Control children are selected from FHSA computerised registers. Twice a year all FHSAs in England and Wales are asked to down-load either their complete registers of children under 15 years of age (as in Trent) or a random sample if this is impractical (as is done in several other health authority regions). Software used during down-loading data to magnetic tape was written for the UKCCS by the FHS Computer Unit. The program arranges records in ascending date of birth order. It can exclude patient's names and addresses, conferring a high degree of anonymity to data files which comprise NHS number, date of birth, sex, date of registration at FHSA, and GP details.

For each case, 10 controls are selected with the same sex, month, and year of birth from the most recent FHSA listing obtained before diagnosis of that case and on which the case appears. The NHS number is used as the only identifier. Every month FHSAs supply the names and addresses of parents or guardians of the controls selected from the lists so that GPs can be approached for permission to contact control families.

#### STUDY C: EARLY POSTMENOPAUSAL INTERVENTIONAL COHORT (EPIC) STUDY

The EPIC study is a randomised, double blind, population based trial of early intervention to prevent bone loss. It compares a new bisphosphonate drug with hormone replacement therapy and placebo. Nottingham is one of four research centres taking part, and the only one in the UK. Each centre has recruited a random sample of 400 local postmenopausal women aged 45 to 59 years. In Nottingham, two stage sampling was employed, based on the Nottinghamshire FHSA register. Firstly, a random sample of 60 general practices was drawn from all those in the Nottingham Health District. The practices were sampled with probabilities in proportion to the number of women aged 45 to 59 on the practice lists supplied by Nottinghamshire FHSA. Senior partners of the randomly selected practices were asked to give permission in principle for their patients to be contacted about the study. At the second stage, for each participating practice, a random sample of approximately 150 women aged 45 to 59 was selected, stratified to obtain approximately equal numbers of postmenopausal women in the three age groups 45–49, 50–54, and 55–59. The large numbers selected from each practice were needed to achieve the total sample size for the trial.

#### Collective experiences in approaching FHSAs for access to registers

Each of the three studies first sought approval of all relevant local ethics committees and this experience will form the basis of a separate report. Applications to use registers were received and vetted by FHSA chief executives, usually in association with their medical and computing advisory teams. Issues of study design, confidentiality, and ethics were subsequently discussed at meetings of interested parties. Factors used by at least one FHSA in screening applications to access registers, as well as those embedded in the Data Protection Act included:

- Whether relevant local or national ethical committee and local medical committee support had been obtained;
- Whether direct patient contact would be involved and, if so, who would make that contact;
- How GPs would be involved, for example, by giving their prior explicit agreement to approaches to individual patients, personal contact with researchers, or being provided with briefing materials;
- Implications for FHSAs in data handling; for example, the range of data required and whether samples or whole down-loads were required. Associated issues of confidentiality were important; for example, security of data, who would hold and access data, and the disposal of listings no longer required; and
- Whether written assurance of handling according to agreed protocols would be given.

#### Access and ease of manipulation of data

In each study a good working relationship was quickly established with all FHSA information departments. All were willing to assist in producing data, although there were delays in doing so in all three studies. These delays were caused firstly by inexperience in procedures requested. For example, one FHSA was unable to provide a sample stratified by age by a jointly agreed date. Secondly, there was a delay in the delivery of specialised software from the FHS Computer Unit at Exeter in two instances. Finally, there was a down-load error in the *Exeter Random Sampling* program. Despite pressure of work several FHSAs volunteered to produce data files using alternative commercial software. These offers were not usually taken up to ensure direct equivalence of data from all sources. Data files were, however, ultimately produced and subsequently imported into *Dbase IV* (Ashton Tate), to allow further manipulation of data.

#### Completeness and accuracy

Checks on the accuracy or completeness, or both, of received information were carried out for all three studies. Different problems and constraints were found in each study.

**STUDY A: TRENT HEALTH LIFESTYLE SURVEY**

Accurate names and addresses were essential as data collection was postal. Inaccuracies in these could lead to failure to contact study subjects. Of the 21 603 addresses in the sample, 2.4% (510) of postcodes were missing and 0.2% (49) were incomplete. The distribution of missing or incomplete postcodes varied between the eight FHSAs (0.1% to 4.8%). Those FHSAs which mainly cover cities had a lower rate of missing or incomplete postcodes (0.1% to 1.0%) than FHSAs with a mainly rural population (2.8% to 4.8%). Overall, 6.4% of all questionnaires were returned as undeliverable, ranging by district from 4.6% to 8.7%. A higher proportion of addresses without or with incomplete postcodes (56 of 559 = 10.0%) than with postcodes (1318 of 21 044 = 6.3%) was returned as undeliverable. Of the 12 338 questionnaires completed, 92% (11 349) of respondents reported both their age and gender in accordance with that indicated by FHSA records with 5% showing lack of accord. The remaining 3% did not report these data.

**STUDY B: UNITED KINGDOM CHILDHOOD CANCER STUDY**

Checks on the completeness of down-loaded data were performed after failure to locate several case children on various Trent region FHSA tapes. It was apparent that not all details of eligible children were being down loaded. Careful audit showed that because of a software error, a systematic omission of whole groups of records on certain dates of birth every month throughout the 15 years of registrations had occurred. Approximately 7% of children were missing from each down-loaded FHSA register, equivalent to missing 1 in 14 eligible children in Trent. This error has now been rectified.

**STUDY C: EPIC TRIAL**

The FHSA first provided details of all relevant general practices and the numbers of women on practice lists aged 45 to 59 years. Of the 60 practices selected, 8% (5) refused to participate. One listed senior partner had retired before the list was produced. During the first 3 months of recruitment, 2593 women were sampled from 26 practice lists. Altogether 0.2% (5) of postcodes were missing from their addresses. Lists of names and addresses of women were sent to their GPs before contacting them. GPs excluded 3% (79) of women for a variety of justifiable reasons, usually based on clinical grounds. Six GPs made a thorough check of names and addresses, although they were not specifically asked to do so. They reported two incorrect names and 11 incorrect addresses.

Letters were sent to the remaining 97% (2514) of women who were asked to indicate on a reply form whether their name and address was correct. A total of 56.4% (1418 of 2514) of all forms were returned. Of these, 4.2% (60)

of the women reported that the title used was incorrect (in most cases Ms or Miss had been used instead of Mrs) and 1.1% (15) reported that the name was incorrect, most being minor errors. Some 5.7% (81) of addresses were reported as being incorrect or incomplete and 2.9% (41) of postcodes were incorrect.

**Discussion**

The projects outlined in this paper together cover survey, case-control, and clinical trial methodology. They focus on various sections of the population: adults aged 16–70, children under 15 years, and women aged 45–59 years and require different data to identify study populations, such as name and address and/or NHS number and/or GP code. Despite the different characteristics of the three studies, permission to use FHSA registers was readily granted after application to FHSA chief executives, usually followed by a visit from senior study personnel. Electoral rolls and telephone and housing listings are more accessible than FHSA registers, being in the public domain, but have the major disadvantage of being less comprehensive and in a printed, rather than computerised format. Computerised electoral rolls are available only in exceptional circumstances.<sup>9</sup> However, software used by FHSAs for sampling caused time consuming difficulties in two of the studies: a general problem in down-loading and the systematic omission of records in the UKCC study. The latter software error might easily have gone unnoticed had a sample rather than a complete list been supplied, as was the practice in other participating regions. All identified software problems have now been solved but highlight the importance of quality checks on data received from FHSAs. For example, checks against an independent register may be made.<sup>10</sup> Alternatively, it is possible to search for obvious systematic omissions by age or name.

FHSA registers cover at least 95% of the total population and cover all ages. Unlike more usual sampling frames they gave all the necessary information for the studies discussed here, such as age, gender, and NHS numbers linking individuals to GPs' records. In addition FHSA lists are more complete than more common sampling frames. Electoral rolls exclude, for example, those under 18 and, of more importance, those who choose not to register,<sup>11</sup> despite non-registration being an offence in law. Non-registration by eligible electors seems to have increased to about 10%,<sup>11</sup> although a pattern is patchy, and it may reach 19% (Riste, MacGregor, Hazes and Silman, personal communication, 1991). FHSA registers are also more complete than telephone listings which are estimated to cover 87% of households in the UK,<sup>3</sup> although 25% of these households are ex-directory.<sup>12</sup> Rural, poorer, less skilled, younger, socially disadvantaged and male individuals have less access to telephones.<sup>12,13</sup> A major problem with telephone listings is that they give no information about household composition and usually list only one household

member. The impact of the de-monopolisation of British Telecom on telephone listings is unclear.

The accuracy of FHSA addresses, including postcodes, was essential for the THLS where GPs were not contacted about individual sample members. Although 6.4% of THLS questionnaires were known to have failed to reach addresses due to errors or omissions in addresses, this percentage for non-delivery is probably an underestimate, as has been found elsewhere.<sup>10,14</sup> Addresses appear to be more inaccurate in inner-city areas and those with high ethnic minority populations.<sup>14</sup> FHSA lists gave most postcodes accurately. These may be linked to other data, such as deprivation scores and ward characteristics, and used with mapping techniques. Envelopes with incomplete (or missing) address or postcode are less likely to reach addresses and both are needed in full to qualify for considerable discounts with the Royal Mail Mailsort Service. Computerised lists with full details of address and postcode can avoid the cost and errors due to retyping or searching for missing data.

FHSA registers are less accurate than newly compiled electoral rolls and telephone listings which are reviewed annually. However, all three lists involve tracking individuals over time and so rapidly become out of date. Only FHSA registers are updated daily. Even so they suffer from time lag before updating. For records of households, such as council tax listings, to be accurate, properties must remain in existence, be occupied, and new housing must be listed. Although more stable than lists of individuals, household registers are less frequently updated. For example, census data listing households are reviewed every 10 years. In Nottingham, new and demolished buildings are notified by the Planning Office to the Council Tax Office every month, although changes reported by the public are registered daily. This pattern may not be universal and researchers may need to check local arrangements.

GP and FHSA records do not always agree, as GP registers are some 3% more up to date.<sup>15,16</sup> Current national projects to link computerised GP and FHSA data should, in the future, overcome lack of accord. In the EPIC study a few participating women reported some inaccuracies although prior screening by some GPs reduced their potential number. In both the EPIC and THLS studies, lack of accord between self reported age and/or gender and FHSA information was low (1%–5%). In the former study the low number of errors was probably explained by the relatively low mobility of women in the age group studied. In the latter it was not possible to say which of the two sources were inaccurate as addressees may not have completed the questionnaire. Another recent study confirms low levels of these type of inaccuracies for example wrong birth dates (5%) and inaccurate names (2%).<sup>15</sup>

### Future developments in the use of FHSA registers

Because of their advantages there will probably be increasing future demand for FHSAs to supply samples or complete lists from their registers for research projects, as an alternative to other less flexible sampling frames. At present there is no unified approach by the FHSAs to these requests. Standardised practice and procedures in submitting and dealing with requests need to be developed by both FHSAs and research workers. These should cover issues such as the efficacy of the proposed research; ethical considerations including confidentiality; practicalities of the supply of data, including timing and checks on the accuracy of software and data; and possible charges by the FHSA for this service. The experience of the three projects has shown that the use of FHSA registers as a sampling frame is not only a viable alternative to the electoral rolls or a telephone listing in all these studies but has considerable advantages over both, particularly in studies of children who are not listed on telephone lists or electoral rolls.

The authors gratefully acknowledge the positive and ongoing support of FHSAs across Trent region for the projects described in this paper. HRR and RD are supported by Trent Regional Health Authority, CACC by Merck, Sharp and Dohme, and CMJ by the UK Co-ordinating Committee on Cancer Research.

- 1 Alderson M. An introduction to epidemiology. London: MacMillan Press, 1983.
- 2 Tones BK, Tilford S and Robinson YK. Health education: effectiveness and efficiency. London: Chapman Hall, 1989.
- 3 Office of Population Censuses and Surveys. (OPCS). General household survey: preliminary results for 1989. London: OPCS, 1990 Monitor SS 90/3.
- 4 Health Promotion Authority for Wales. Welsh heart health survey. 1985. Protocol and questionnaire with preliminary report of the Welsh heart health survey. Cardiff: Health Promotion Authority for Wales, 1985. Heartbeat Wales Technical report 2; 4. HPAW.
- 5 Page H. The use of the FHSA registration index for a health needs survey. *Health Services Management Research* 1991; 4:140–7.
- 6 Nicoll A, Bassett K, Ulijaszek SJ. What's in a name? Accuracy of using surnames and forenames in ascribing Asian ethnic identity in English populations. *J Epidemiol Community Health* 1986;40:364–8.
- 7 Jacoby A. Possible factors affecting response to postal questionnaires findings from a study of general practitioner services. *J Public Health Med* 1990;12:131–5.
- 8 Roberts HR, Dengler R. Trent health lifestyle survey interim report, 1991/92. Nottingham: Department of Public Health Medicine and Epidemiology, University of Nottingham, 1992.
- 9 Galt MP. Health-related attitudes and behaviours in 18/19 year olds. A comparison of data collection methods. Nottingham: Department of Public Health Medicine and Epidemiology, University of Nottingham, 1991. Dissertation.
- 10 Bickler G, Sutton S. Inaccuracy of FHSA registers: help from electoral registers. *BMJ* 1993;306:1167.
- 11 Hickman M. *Compiling the electoral register*. London: HMSO, 1990. OPCS Social Surveys Division, SS;1280B.
- 12 Collins M, Sykes W. The problems of non-coverage and unlisted number in telephone surveys in Britain. *J R Stat Soc* 1987;150:241–53.
- 13 Catlin G, Shields M. Conducting community based studies by telephone. *Can J Public Health* 1988;79:29–31.
- 14 Doyle Y. A survey of the cervical screening service in a London District. *Soc Sci Med* 1991;32:953–7.
- 15 Voss SN, Thomas HC. How well do family practitioner committee and general practice records agree? Experience in a semi-rural practice. *Br J Gen Pract* 1991;41:293–4.
- 16 Fraser RC, Clayton PG. The accuracy of age/sex registers: practice medical records and the family practitioner committee registers. *J R Coll Gen Pract* 1981;31:410–9.